

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 October 2003 (02.10.2003)

PCT

(10) International Publication Number
WO 03/081523 A1

(51) International Patent Classification⁷: G06K 9/64, 9/00, G06T 7/00

(21) International Application Number: PCT/GB03/01209

(22) International Filing Date: 21 March 2003 (21.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02252097.7 22 March 2002 (22.03.2002) EP

(71) Applicant (for all designated States except US): **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **STENTIFORD, Frederick, Warwick, Michael** [GB/GB]; Sheepstor, Boyton, Woodbridge, Ipswich, Suffolk IP12 3LH (GB).

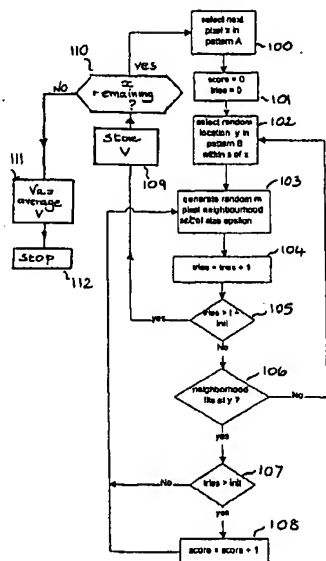
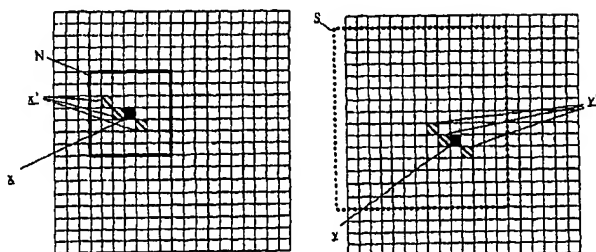
(74) Agent: **LLOYD, Barry, George, William**; BT GROUP LEGAL INTELLECTUAL PROPERTY DEPARTMENT, HOLBORN CENTRE, 8th Floor, 120 Holborn, London EC1N 2TE (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: COMPARING PATTERNS



(57) Abstract: A first image (or other pattern) is represented by a first ordered set of elements A each having a value and a second pattern is represented by a second such set. A comparison of the two involves performing, for each of a plurality of elements x of the first ordered set the steps of selecting (103) from the first ordered set a plurality of elements x' in the vicinity of the element x under consideration, selecting (102) an element y of the second ordered set and comparing the elements x' of the first ordered set with elements y' of the second ordered set (each of which has the same position relative to the selected element y' of the second ordered set as a respective one x' of the selected plurality of elements of the first ordered set has relative to the element x under consideration). The comparison itself 106 comprises comparing the value of each of the selected plurality of elements x' of the first set with the value of the correspondingly positioned element y' of the like plurality of elements of the second set in accordance with a predetermined match criterion to produce a decision that the plurality of elements of the first ordered set matches the plurality of elements of the second ordered set. The comparison is then repeated with a fresh selection of the plurality of elements x' of the first set and/or a fresh selection of an element y of the second ordered set generating a similarity measure V as a function of the number of matches. Preferably, following a comparison resulting in a match decision, the next comparison is performed with a fresh selection of the plurality of elements x' of the first set and the same selection of an element y of the second set.



Published:

— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Comparing Patterns

The present invention is concerned with comparing one pattern with another, and is of particular interest in the comparison of two-dimensional patterns such as visual images, although
5 applicable also to one-dimensional patterns and patterns having three or more dimensions. Standard approaches to pattern recognition use templates to recognise and categorise patterns [1]. Such templates take many forms but they are normally produced by a statistical analysis of training data and matched with unseen data using a similarity measure [2]. The statistical analysis is normally carried out over a number of intuitively selected features that appear to satisfy the needs of the
10 recognition task. For example, in speech recognition templates can be encapsulated as Hidden Markov Models derived in the frequency domain and in Optical Character Recognition the templates take the form of the character fonts themselves. In the case of face recognition a number of intuitively chosen features such as skin texture, skin colour and facial feature registration are used to define face templates. In a CCTV surveillance application intruders are normally detected
15 through a process of frame subtraction and background template modelling which detects movement and removes background effects from the processing [3]. In many cases the number of features leads to a computationally unmanageable process and Principal Components Analysis and other techniques are used to scale down the problem without significantly reducing performance [<http://www.partek.com/index.html>]. These approaches achieve great success in non-noisy
20 environments but fail when the pattern variability and number of pattern classes increase.

Some techniques for analysis of images or other patterns where the pattern is compared with other parts of the same pattern are described in our earlier patent applications as follows.
European patent application 00301262.2 (publication No. 1126411) (applicants ref. A25904EP#);
International patent application PCT/GB01/00504 (publication No. WO 01/61648) (applicants ref.
25 A25904WO);
International patent application PCT/GB01/03802 (publication No. WO02/21446) (applicants ref. A25055WO);
U.S patent application 977,263/09 filed 16 October 2001 (publication No. 20020081033) (applicants ref. A25904US1);
30 - as well as the following papers published by the inventor:

Stentiford F W M, "An estimator for visual attention through competitive novelty with application to image compression", Proc. Picture Coding Symposium 2001, Seoul, 25 – 27 April, pp 101 - 104, 2001.

Stentiford F W M, "An evolutionary programming approach to the simulation of visual attention", Proc. Congress on Evolutionary Computation 2001, Seoul, pp 851 - 858, 27 - 30 May, 2001.

According to one aspect of the present invention there is provided a method of comparing a first pattern represented by a first ordered set of elements each having a value with a second pattern
5 represented by a second ordered set of element each having a value, comprising performing, for each of a plurality of elements of the first ordered set the steps of:

(i) selecting from the first ordered set a plurality of elements in the vicinity of the element under consideration, the selected elements having, within the ordered set, respective mutually different positions relative to the element under consideration;

10 (ii) selecting an element of the second ordered set;

(iii) comparing the selected plurality of elements of the first ordered set with a like plurality of elements of the second ordered set each of which has within the second ordered set the same position relative to the selected element of the second ordered set as a respective one of the selected plurality of elements of the first ordered set has relative to the element under
15 consideration, said comparison comprising comparing the value of each of the selected plurality of elements of the first ordered set with the value of the correspondingly positioned element of the like plurality of elements of the second ordered set in accordance with a predetermined match criterion to produce a decision that the plurality of elements of the first ordered set matches the plurality of elements of the second ordered set;

20 (iv) repeating said comparison with a fresh selection of the plurality of elements of the first ordered set and/or a fresh selection of an element of the second ordered set; and

(v) generating for the element under consideration a similarity measure as a function of the number of comparisons for which the comparison indicates a match.

Other aspects of the invention are defined in the claims.

25 Some embodiments of the present invention will now be described with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of an apparatus for performing the invention;

Figure 2 is a diagram illustrating operation of the invention;

Figure 3 is a flowchart of the steps to be performed by the apparatus of Figure 1 in
30 accordance with one embodiment of the invention; and

Figures 4 to 9 illustrates some images and numerical results obtained for them

- 3 -

Figure 1 shows an apparatus consisting of a general purpose computer programmed to perform image analysis according to a first embodiment of the invention. It has a bus 1, to which are connected a central processing unit 2, a visual display 3, a keyboard 4, a scanner 5 (or other device, not shown) for input of images, and a memory 6.

5 In the memory 6 are stored an operating system 601, a program 602 for performing the image analysis, and storage areas 603, 604 for storing two images, referred to as image A and image B. Each image is stored as a two-dimensional array of values, each value representing the brightness of a picture element within the array.

The image arrays are shown schematically in Figures 2a and 2b. Image A consists of a 20
10 x 20 array of picture elements $\underline{x} = (x_1, x_2)$ where x_1 and x_2 are the horizontal and vertical positions of the elements within the image. Each element has a brightness represented by a respective value $a = a(\underline{x})$. Similarly, the image B consists of a 20 x 20 array of picture elements $\underline{y} = (y_1, y_2)$ having brightness values b .

A method of image analysis performed by the program 602 is performed in accordance
15 with the flowchart shown in Figure 3. It has as its aim to generate a measure V of similarity between the images; more specifically, a high value of V indicates that image B contains visual material similar to the content of image A.

It is assumed that the images have width x_{max_1} and y_{max_1} and height x_{max_2} and y_{max_2} , and the picture element co-ordinates are $x_1 = 0 \dots x_{max_1}-1$, $x_2 = 0 \dots x_{max_2}-1$.

20 In Step 100, a picture element \underline{x} in image A is selected. They can be dealt with in any order, but it is convenient to select initially the element $\underline{x} = (\epsilon, \epsilon)$ and on subsequent iterations, successive elements in raster-scan fashion, until all have been dealt with, except for those within ϵ of the edge of the image i.e. up to $(x_{max_1}-\epsilon-1, x_{max_2}-\epsilon-1)$.

In Step 101, the score V and a counter *tries* are set to zero.

25 In Step 102, an element $\underline{y} = (y_1, y_2)$ is chosen at random from image B, within a maximum distance s of the position of \underline{x} , that is

$$|x_i - y_i| \leq s \quad \text{for all } i$$

This limitation, which is optional, is represented in Figure 2b by the dotted square S . It improves efficiency if it is known that corresponding objects in the two images will not be shifted
30 in position by more than s picture elements in either co-ordinate direction: in effect, s represents the maximum misregistration or local distortion between the two images.

- 4 -

The selection of \underline{y} is also subject to the restriction that it is located no more than ε from the edge of the image, i.e.

$$\varepsilon \leq y_i \leq y_{\max_i} - \varepsilon - 1 \quad \text{for } i = 1, 2$$

where y_{\max_i} is the width or height of the image in picture elements (20 in this example).

- 5 In Step 103, a set S_x of picture elements \underline{x}' , are selected, comparing the element plus at least one further element, m in number, randomly selected from image A, in the neighbourhood N of picture element \underline{x} , where the neighbourhood N is a $2\varepsilon+1 \times 2\varepsilon+1$ square centred on \underline{x} . That is, an element $\underline{x}' = (x'_1, x'_2)$ lies within N if

$$|x'_i - x_i| < \varepsilon \quad \text{for all } i$$

- 10 We prefer that the neighbourhood set S_x includes the element \underline{x} itself, although this is not actually essential. A neighbourhood set S_y is then defined in image B, with $m+1$ elements $\underline{y}' = (y'_1, y'_2)$ each of which has the same position relative to element \underline{y} as the corresponding element \underline{x}' of S_x has from \underline{x} , that is

$$y'_i - y_i = x'_i - x_i \quad \text{for all } i$$

- 15 (The reason for the edge restrictions mentioned above is of course to avoid choosing a neighbourhood which extends outside the image area). In Step 104, the counter *tries* is incremented, and in Step 105 its value is checked so see whether sufficient comparisons have been performed.

- Assuming for the present that this is not the case, then at Step 106 the two sets S_x, S_y are
20 compared. If the difference between the value of an element \underline{x}' of the set S_x and the value of the correspondingly positioned element \underline{y}' of the set S_y is less than a threshold δ , i.e.

$$|a(\underline{x}') - b(\underline{y}')| < \delta$$

then the two elements are considered to match. The two sets are considered to match only if every element \underline{x}' of the set S_x matches the corresponding element \underline{y}' of the set S_y .

- 25 If the sets do not match, the element \underline{y} is discarded as a candidate for further comparisons and the process resumes at step 102 with a fresh point \underline{y} being chosen. If on the other hand, they match, then \underline{y} is retained for a further iteration. Ignoring Step 107, for the present, the process proceeds with step 108 where the score V is incremented. Different neighbourhood sets are then generated at Step 103 and the process repeated from that point.

- 5 -

When, at Step 105, the count *tries* of the number of comparisons exceeds a threshold, the process for the current \underline{x} is terminated: at Step 109 the value of V is stored in the memory 6 and then, at step 110, a check is performed as to whether all points \underline{x} have been dealt with. If so, then firstly the average score V_a (being the average of all the individual scores V for the picture elements \underline{x} in image A) is calculated at Step 111 and the process terminated at 112; otherwise the whole process is repeated from Step 100 for a fresh point \underline{x} .

The threshold value used at Step 105 could be set to a desired number of comparisons t , and Step 107 omitted. In practice, however, we find that the early comparisons often do not find a match, but once a match is found a significant number of different neighbourhood sets show matches with the same \underline{y} . In the interests of providing a score V which is easily interpreted, the first few iterations are not scored; the process as shown includes a test 107 which prevents the score being incremented during the first *init* comparisons. The threshold used at Step 105 is thus increased by *init* so that the maximum potential score is t . In this way, we find that the chances of obtaining a high or maximum score for images which are very similar or identical is greatly increased.

Specific values for the various parameter used in tests were as follows:

Image A size $x_{max_1} \times x_{max_2}$: 20×20
 Image B size $x_{max_1} \times x_{max_2}$: 20×20
 Neighbourhood parameter ϵ : 4
 Neighbourhood set size m : 3
 Maximum A - B shift s : 7
 Number of comparisons t : 50
 Initial comparisons *init*: 20

Naturally, for higher definition images, correspondingly larger values of ϵ and s would be chosen. The values of t and *init* have to be increased as s increases in order to obtain equally reliable and statistically significant results.

The results of these tests are given later; where the parameters used differ from those quoted above, this is noted.

It will be seen that the score of a picture element \underline{x} in the pattern A is high if many randomly selected S_x match S_y for a given \underline{y} in pattern B. It is worth noting that the Visual Attention score (discussed in our earlier patent applications) of picture element \underline{x} in pattern A is high if S_x mis-matches S_y where \underline{y} is *not* held but is randomly selected from pattern A. Whereas

- 6 -

Visual Attention requires no memory save the single image A, the present method requires also the information contained in pattern B to detect commonality.

A location \underline{x} will be worthy of cognitive attention if a sequence of t neighbourhood sets S_x matches a high proportion of the same neighbourhoods surrounding some \underline{y} in pattern B. In Figure
 5 2, $m = 3$ picture elements \underline{x}' are selected in the neighbourhood of a picture element \underline{x} in pattern A and matched with 3 picture elements in the neighbourhood of picture element \underline{y} in pattern B. Each of the picture elements might possess three colour intensities, so $\underline{a} = (a_r, a_g, a_b)$ and the neighbourhood set of the second picture element \underline{y} matches the first if the colour intensities of all $m + 1$ corresponding picture element have values within δ of each other. Picture elements \underline{x} in A that
 10 achieve large numbers of matches over a range of t neighbourhood sets S_x with neighbourhood sets S_y around \underline{y} in B are assigned a high score. This means that neighbourhood sets in A possessing structure present in B will be assigned high scores.

Some image analysis techniques carry out comparison calculations between images using patches that are neighbourhoods in which all the picture elements are employed. Patches match
 15 when a measure of correlation exceeds a certain threshold. This approach is unable to make best use of detail that is smaller than the size of the patch except in the case in which the correlation measure is designed to identify a specific texture. The random picture element neighbourhoods S_x used in this method do not suffer from this disadvantage.

The gain of the scoring mechanism is increased significantly by retaining the picture
 20 element location \underline{y} if a match is detected, and re-using \underline{y} for comparison with the next of the t neighbourhood sets. It is likely that if a matching picture element configuration is generated, other configurations will match again at the same point, and this location \underline{y} once found and re-used, will accelerate the rise of the score provided that the sequence is not subsequently interrupted by a mismatch.

25 If, however, S_x subsequently mis-matches at that location, the score is not incremented, and an entirely new location \underline{y} in pattern B is randomly selected ready for the next comparison. In this way competing locations in pattern B are selected against if they contain little commonality with the neighbourhood of \underline{x} in pattern A.

Some possible variations will now be discussed.

30 a) It is not essential that all picture elements \underline{x} be analysed: thus, if desired, a subset – perhaps on a regular grid – could be chosen.

b) The above method assumes single (brightness) values for the picture elements. In the case of colour images, the images could firstly be converted to monochrome. Alternatively, if

- 7 -

each picture element is expressed as three values, such as red, green and blue, or luminance plus colour difference, then the test for a match would involve consideration of the distance between the element values in 3 dimensions. For example, if the colour component values are $\underline{a} = (a_r, a_g, a_b)$, $\underline{b} = (b_r, b_g, b_b)$ for the two image then the criterion for a match might be $(a_r - b_r) < \delta$ and $|a_g - b_g| < \delta$ and $|a_b - b_b| < \delta$. See below for further discussion of distance measures.

c) The strategy of choosing \underline{y} at random could be modified by providing that when a element $\underline{x} = (x_1, x_2)$ produces a high score V against element \underline{y} and an adjacent element \underline{x} (e.g. $(x_1 + 1, x_2)$) is to be dealt with, then the first element \underline{y} to be selected in image B could be the correspondingly adjacent element (i.e. $(y_1 + 1, y_2)$). Further selections of \underline{y} would be random, as before. In the event that the high score for element \underline{x} was produced as a result of comparisons with more than one \underline{y} , then one would choose the element adjacent to whichever \underline{y} made the largest contribution to the high score. To implement this it would be necessary to track, temporarily, the coordinates of the different picture elements \underline{y} and their partial scores.

d) The process need not necessarily be carried out for the whole image. For example, if a region of image A has been identified as being of special interest – perhaps using the method described in one of our earlier patent applications referred to above – then the picture elements \underline{x} dealt with may be just those lying within the identified region.

References here to random selection include the possibility of selection using a pseudo-random process.

The method is not applicable solely to images, or even three-dimensional patterns, but can be used for any ordered set of values in 1, 2, 3 or more dimensions.

More generally, suppose that one wishes to generate a measure of the similarity between two patterns A and B, where both the patterns and the individual pattern of values within a pattern may have 1, 2 or more dimensions.

Let a set of measurements \underline{a} on pattern A correspond to a location \underline{x} in A in bounded n-space $(x_1, x_2, x_3, \dots, x_n)$ where

$$\underline{x} = (x_1, x_2, x_3, \dots, x_n) \text{ and } \underline{a} = (a_1, a_2, a_3, \dots, a_p)$$

Define a function \underline{F} such that $\underline{a} = \underline{F}(\underline{x})$ wherever \underline{a} exists. It is important to note that no assumptions are made about the nature of \underline{F} e.g. continuity. It is assumed that \underline{x} exists if \underline{a} exists.

Consider a neighbourhood N of \underline{x} where

$$\{\underline{x}' \in N \text{ iff } |x_i - x'_i| < \epsilon_i \forall i\}$$

- 8 -

Select a set of $m + 1$ random points S_x in N where

$S_x = \{\underline{x}'_1, \underline{x}'_2, \underline{x}'_3, \dots, \underline{x}'_m\}$ and $\underline{F}(\underline{x}'_i)$ is defined.

- where however one of the \underline{x}' is preferably constrained to equal \underline{x} . It is not found in practice to be necessary to constrain the random selection to avoid selection of the same point
5 twice: i.e. the \underline{x}'_i need not be distinct. Indeed, the \underline{x}'_i could all be coincident with \underline{x} or \underline{x}'_1 .

Select a location \underline{y} corresponding to the set of measurements \underline{b} on pattern B for which \underline{F} is defined.

Define the set $S_y = \{\underline{y}'_1, \underline{y}'_2, \underline{y}'_3, \dots, \underline{y}'_m\}$ where

$\underline{x} - \underline{x}'_i = \underline{y} - \underline{y}'_i$ and $\underline{F}(\underline{y}'_i)$ exists.

10 The neighbourhood set S_x of \underline{x} is said to match that of \underline{y} if

$$|F_j(\underline{x}) - F_j(\underline{y})| < \delta_j \text{ and } |F_j(\underline{x}'_i) - F_j(\underline{y}'_i)| < \delta_j \quad \forall i, j.$$

In general δ_j is not a constant and will be dependent upon the measurements under comparison i.e.

$$\delta_j = f_j(\underline{F}(\underline{x}), \underline{F}(\underline{y})) \quad j = 1 \dots p$$

15 Note that for some problems rather than defining a match as requiring the individual components $a_j = F_j(\underline{a})$ and $b_j = F_j(\underline{b})$ to match one might prefer to apply a threshold to the vector distance between \underline{a} and \underline{b} such as the city block distance.

$$\sum_j |a_j - b_j|$$

or the Euclidean distance

20

$$\sqrt{\left\{ \sum_j (a_j - b_j)^2 \right\}}$$

The operation of the method described in the flowchart of Figure 3 will be illustrated by the following test results.

In the tests, 20 x 20 black and white images were used (i.e. the values of a and b were
25 always either 0 or 1). The tests however differ from the earlier description in that only black picture elements are scored – that is, the match criterion is not merely that $|a-b| < \delta$ but that $a=b=1$.

In each case the accompanying Figure shows a first image A, plus several second images B with which the first was compared, marked B1, B2 etc. (in once case two first images A1, A2 are compared with one image B) . The coordinates x_1, y_1 are measured from left to right and the coordinates x_2, y_2 from top to bottom. Each of the second images is marked with the average score

5 V_a . Some Figures show a three-dimensional plot of the individual scores V over the image area: in each case the right-hand axis represents the horizontal axis of the mark and the front axis represents the vertical axis of the image. The vertical axis of the plot represents the value of V (in the embedded recognition case of Example 2, and $50-V$ in the other cases).

Example 1. *Optical Character Recognition (Figure 4)*

10 An image A of the letter B is compared with images B1 – B5, of the letters A, B, C, D and E. The rank ordered average scores V_a correspond to B, D, E, C, A. The results were generated with $t=50, s=3$ and $\epsilon = 2$.

Example 2. *Embedded Recognition (Figure 5)*

An image A of an 'X' shape is compared with another image B1 containing 9 smaller shapes including an 'X'. The 'X' shape is again compared with image B2 with the same 9 shapes except for the small 'X' which is altered. The average scores V_a are 27.5 and 11.9 respectively, indicating the presence of 'X' features in the first and less in the second. The individual scores V for the respective comparisons are shown in the charts. The results were generated with $t=50, s=7$ and $\epsilon = 2$. A more complicated example of embedded recognition is the identification of known

20 faces in a crowd of people.

Example 3. *Trade mark Retrieval (Figure 6)*

An image A of a special cross shape is first compared with 4 basic shapes B1-B4 including a large cross, each basic pattern representing clusters of similar shapes with high scores between each other. The large cross generates the largest average score V_a (43.6). The same special cross

25 in then compared with 4 varieties of cross B5-B8 including a small version of itself (B5) which obtains a maximum score of 50. The chart provides the values of $(50 - V)$ for each picture element in the comparison with the large cross – the tall columns indicate the disparity with the ends of the lower legs of the cross. The results were generated with $t=50, s=7$ and $\epsilon = 2$.

Example 4. *Fingerprint Recognition (Figure 7)*

30 More complex patterns, some possessing a great deal of similar material, are able to be distinguished and retrieved as in the trade mark example. A simplified arch A matches best with a representative arch B4 (49.3), and then with an identical pattern (49.9) amongst other arches. The

chart provides the values of $(50 - V)$ for each picture element in the comparison with the representative arch pattern – the tall columns indicate the disparity with the ridge ends and on the border. The results were generated with $t=50$, $s=3$ and $\varepsilon = 2$.

Example 5. Face Recognition (Figure 8)

5 Although the data is extremely simplified, this example illustrates an implementation of face recognition. A face A is matched with 4 representative faces B1 – B4 and then with 4 faces in a sub-cluster B5 – B8 represented by the best fitting representative face B2 (45.7). The best fit is with a non-identical face B7 in the sub-cluster (49.6). The chart provides the values of $(50 - V)$ for each picture element in the comparison with the 2nd closest pattern B6 (48.2) – the tall columns
10 indicate the disparity in the mouth region. The results were generated with $t=50$, $s=3$ and $\varepsilon = 2$.

Example 6. Surveillance (Figure 9)

This example illustrates how intruders might be detected against a variable background. A pattern A1 representing a moving cloud, tree and fence obtains a high average score V_a (48.0) against a standard background B, but a version (A2) with an intruder present obtains a much lower
15 score (33.1). The chart provides the values of $(50 - V)$ for each picture element in the intruder image (33.1) – the tallest columns indicate the location of the intruder. The results were generated with $t=50$, $s=3$ and $\varepsilon = 2$.

This method carries out an analysis of two images that identifies perceptually significant common features and differences without being affected by scale differences, and by local relative
20 distortions and translations. This means that it can be applied to recognition problems in which the object to be recognised is embedded in another image. It also means that images may be categorised according to the measure of feature similarity with each other and these values used to define a multi-class recognition system and a Query By Example retrieval system. By the same token the method may be used to detect dissimilarities as in the case of intruder detection or forgery
25 detection. It may be used to detect disparities between images to detect motion or parallax. It does not necessarily rely upon prior knowledge of the content of the images and no training process is necessary. The algorithm is eminently suitable for parallel implementation.

Various advantages of the methods we have described will now be discussed.

Pattern Recognition

30 Standard approaches to pattern recognition require large numbers of representative patterns from all classes in order to satisfy the requirements of the training process for the classifier. Such classifiers will still fail if the features selected do not characterise unseen patterns not represented

- 11 -

properly in the training set. This method requires no training set, save the two images under comparison.

It is normal for pattern recognition techniques to make special provision for scale differences between the patterns being classified. Such differences in the visual domain can arise because of perspective and other factors and it is common for prior knowledge to be incorporated in the classifier to compensate. This is not necessary in this method, provided that ϵ is not too large.

Another advantage is the ability to identify patterns embedded within larger data structures without suffering from exponentially growing computational demands or the effects of noise. This method therefore has application to the problem of detecting copyright infringement where portions of material have been cropped from larger works (of art, for example), and the task of reducing the size of databases where it is known that duplication is prevalent. It also has direct application to the problem of identifying scene-of-crime fingerprints where only a part of the total print is available for matching. In the case of facial recognition the method lends itself to searches based upon restricted portions of the unknown face. This means, for example, that searches can be carried out purely on the basis of the eye and nose region in cases where beards and moustaches might lead to ambiguity.

Unless provision is made for specific pattern content, the standard template approaches to pattern recognition fail when the patterns under comparison differ because of local distortions or small movements, as would be the visual case with rustling trees, moving clouds, changes in facial expression, scene-of-crime fingerprints on irregular surfaces, or noise, for example. Such provision requires prior knowledge of the application and will still cause the system to fail if the unseen pattern distortions do not conform to the system design requirements. This method is able to ignore the effects of local distortions without prior knowledge of the type of distortion.

The method may be applied to patterns of any dimension, such as one-dimensional audio signals, three dimensional video data (x,y,time), or n-dimensional time dependent vectors derived from any source such as sensor arrays. In the case of speech recognition it is able to handle variations in the speed of the speech without the use of special heuristics. A conventional approach uses Dynamic Time Warping to overcome this problem, but invokes greater computational effort and the danger of the warping process leading to increased misclassifications especially in a large multi-class problem. Furthermore portions of utterances would be sufficient for word identification using this invention if they were unique in the domain of discourse (e.g. 'yeah' instead of 'yes', missing the final sibilant).

This method deals with these problems by measuring the score of neighbourhoods in each image. Whereas a picture element is important within a single image if its neighbourhood is significantly different from most other parts of the same image (as discussed in our earlier patent applications mentioned above), it is *cognitively* important if its neighbourhood is significantly
5 similar to that in a neighbourhood in the second image. The effects of mis-registration and local distortions are considerably reduced as the scores are not affected within certain limits.

The scores can be used to cluster groups of patterns possessing high scores relative to each other. Representative patterns taken from each cluster may themselves be clustered to form super-clusters and the process continued to structure very large pattern databases. Query-By-Example
10 retrieval can be carried out by measuring the scores to each of the top level representative patterns and then to each of the representative patterns in sub-clusters corresponding to the highest scoring representative pattern in the previous cluster. It is likely that some representative patterns will represent overlapping clusters containing patterns which happen to possess high scores with those representative patterns. The method may be used in combination with a visual attention mechanism
15 to rapidly define an area of interest in the pattern and then derive scores between this restricted area and a set of reference patterns (B patterns). In an aviation context a flying object might be quickly detected using visual attention and subsequently computed scores would reveal whether the object was likely to be a bird or a plane.

Disparity Detection

20 Standard approaches to disparity detection rely heavily upon accurate registration between two images so that the subtraction (which may be carried out piecewise for a small areas of the total image) takes place between picture elements that correspond to the same points on the original object pictured in the two images. The resulting difference-image highlights those areas that correspond to differences in the original images. This becomes extremely difficult if noise is
25 present as uncertainty is introduced into the estimate of the correct registration position and many spurious differences can be generated as a result.

Even if noise is absent local distortions or slight subject movements will cause mis-registration and areas of difference will be highlighted which are of little interest unless the distortion or movement itself is being measured. Linear or non-linear digital image registration
30 techniques prior to subtraction partially compensates but does not eliminate this problem in a large proportion of cases [4].

A severe problem faced by conventional methods arises from the different conditions under which the two images were created. Such differences may stem from the lighting, weather

- 13 -

conditions, a slightly different viewpoint, a different make of film, or a different camera with a different optical system. These interferences all contribute towards the generation of spurious differences between the images that do not relate to the content.

5 Differences between the two images caused by the conditions of image capture will be factored out by this method because such differences will affect the score equally across all parts of the image and will not disturb the rank ordering of the individual scores. Significant illumination differences between the two images are compensated by employing a larger threshold for the picture element matching (δ_j see above). This would be necessary, for example, if X-ray exposure times were different between the two images.

10 This method also has the advantage of being able to detect multiple disparities in which image A is compared with images B1,B2,B3 etc. This would be useful in the case of CCTV intruder detection in which image frames B1,B2,B3 etc would be typical examples of different atmospheric conditions and other normal background states, and an alarm would be raised only if disparities were detected in all the normal image frames.

15 It will be seen that these methods have relevance to almost all applications requiring pattern recognition:

Face recognition, fingerprint recognition, OCR, image retrieval, trademark identification, forgery detection, surveillance, medical diagnosis, and others.

20 They are particularly relevant to the medical industry where the comparison of very variable and noisy images is necessary to track the progress of medical conditions. For example, the identification of likely cancerous growth in mammograms is often gleaned from X-rays taken at different times and any automatic assistance would increase radiologist's throughput. In another example, contrast enhancement is normally obtained by subtracting the X-ray image of the normal breast from one taken after contrast injection and obtaining proper registration is not only difficult
25 to obtain but is critical to the success of the process. The methods we have described can provide a clearer result without the attendant registration problems.

Advanced software detection systems for mammograms (eg. <http://www.r2tech.com/prd/>) are quite capable of spotting specific features commonly associated with cancer (eg microcalcifications), but the problem of temporal comparisons is largely not addressed.

30 References

[1] Vailaya A et al., Image Classification for Content-Based Indexing, IEEE Trans on Image Processing, Vol 10, No 1, pp 117 – 130, Jan 2001.

- 14 -

[2] Santini S & Jain R, Similarity Matching, in Proc 2nd Asian Conf on Computer Vision, pages II 544-548, IEEE, 1995.

[3] IEEE Trans PAMI – Special Section on Video Surveillance, vol 22 No 8, Aug 2000.

[4] Brown L G, A survey of image registration techniques, ACM Computing Surveys,
5 Vol.24, No. 4 (Dec. 1992), pp. 325-376.

CLAIMS

1. A method of comparing a first pattern represented by a first ordered set of elements each having a value with a second pattern represented by a second ordered set of element each having a value, comprising performing, for each of a plurality of elements of the first ordered set the steps of:

(i) selecting from the first ordered set a plurality of elements in the vicinity of the element under consideration, the selected elements having, within the ordered set, respective mutually different positions relative to the element under consideration;

10 (ii) selecting an element of the second ordered set;

(iii) comparing the selected plurality of elements of the first ordered set with a like plurality of elements of the second ordered set each of which has within the second ordered set the same position relative to the selected element of the second ordered set as a respective one of the selected plurality of elements of the first ordered set has relative to the element under consideration, said comparison comprising comparing the value of each of the selected plurality of elements of the first ordered set with the value of the correspondingly positioned element of the like plurality of elements of the second ordered set in accordance with a predetermined match criterion to produce a decision that the plurality of elements of the first ordered set matches the plurality of elements of the second ordered set;

20 (iv) repeating said comparison with a fresh selection of the plurality of elements of the first ordered set and/or a fresh selection of an element of the second ordered set; and

(v) generating for the element under consideration a similarity measure as a function of the number of comparisons for which the comparison indicates a match.

25 2. A method according to claim 1 in which the plurality of elements of the first ordered set includes the element under consideration.

3. A method according to claim 1 or 2 in which the selection of the plurality of elements of the first ordered set is random or pseudo-random.

- 16 -

4. A method according to claim 1,2 or 3 in which the selection of an element of the second ordered set is random or pseudo-random.
5. A method according to any one of the preceding claims in which, following a comparison resulting in a match decision, a further comparison is performed with a fresh selection of the plurality of elements of the first ordered set and the same selection of an element of the second ordered set.
6. A method according to any one of the preceding claims in which, following a comparison resulting in a decision that there is no match, a further comparison is performed with a fresh selection of an element of the second ordered set and the same selection of the plurality of elements of the first ordered set.
7. A method according to any one of the preceding claims, in which each value comprises a plurality of components.
8. A method according to any one of the preceding claims, in which the comparison is performed a predetermined number of times.
9. A method according to any one of the preceding claims, in which the similarity measure is a function of the number of comparisons for which the comparison indicates a match, excluding an initial number of comparisons.
10. A method according to any one of the preceding claims in which, when an element under consideration has resulted in a similarity measure indicative of a large number of matches and a further element of the first ordered set is to be considered, the first selection of an element of the second ordered set for comparison therewith is of that element of the second ordered set whose position, within the second ordered set, relative to an element of the second ordered set giving rise to said large number of matches, is the same as the position of the further element of the first ordered set relative to the previous element under consideration.

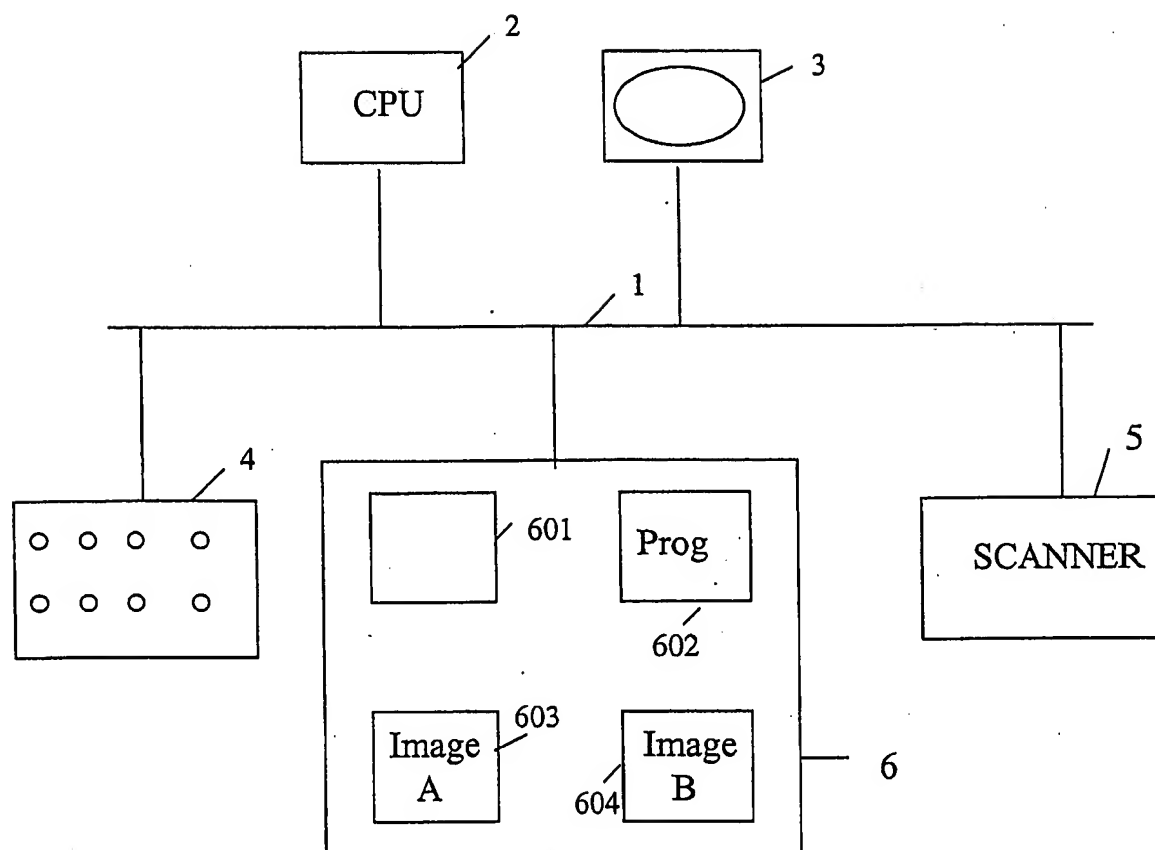


Figure 1

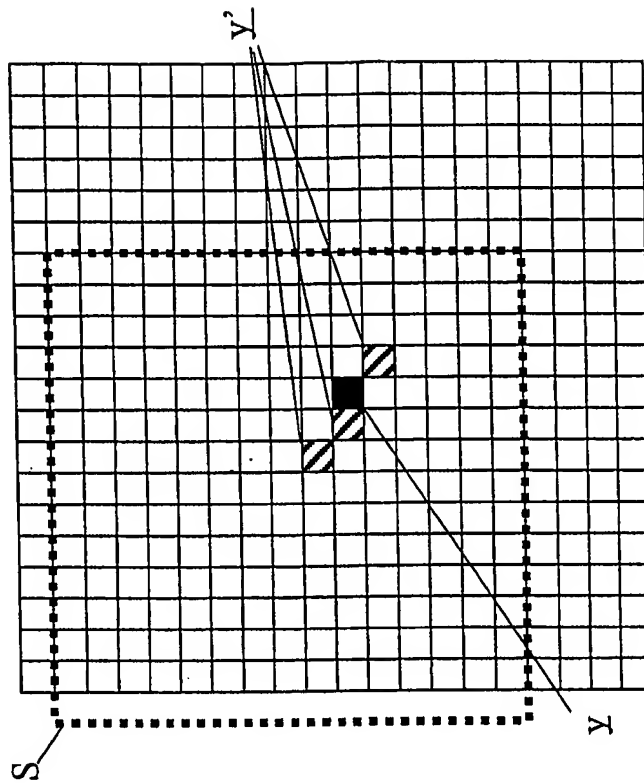


Figure 2b

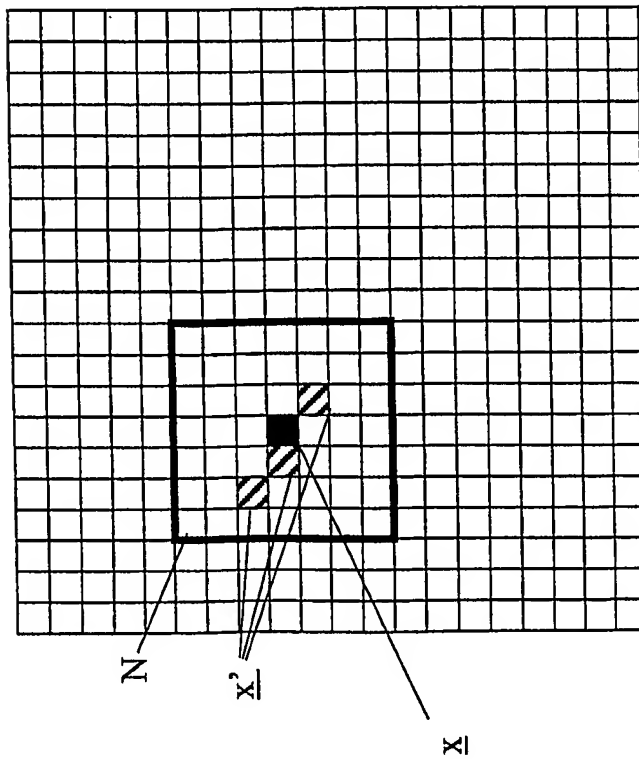


Figure 2a

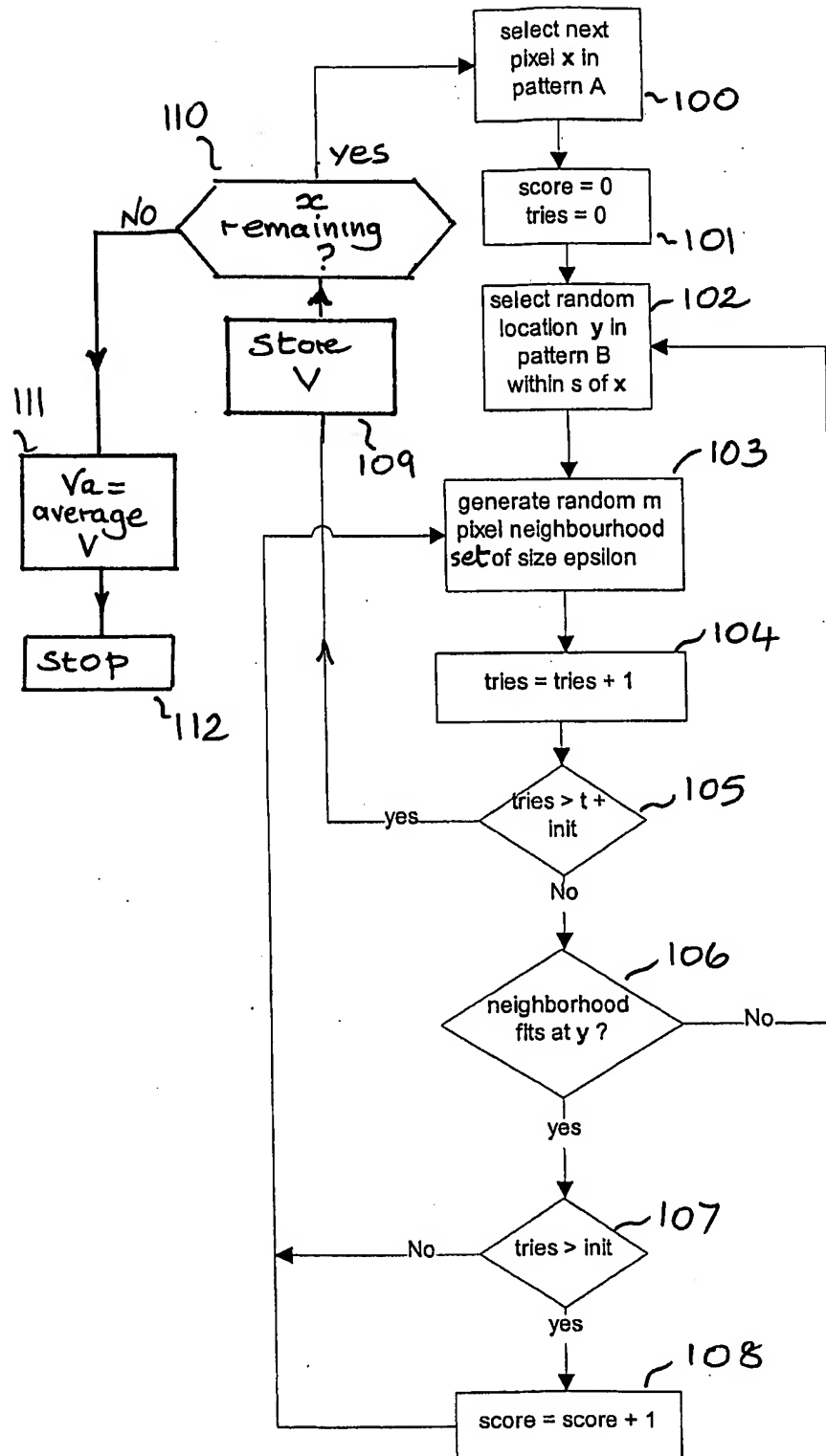


Figure 3

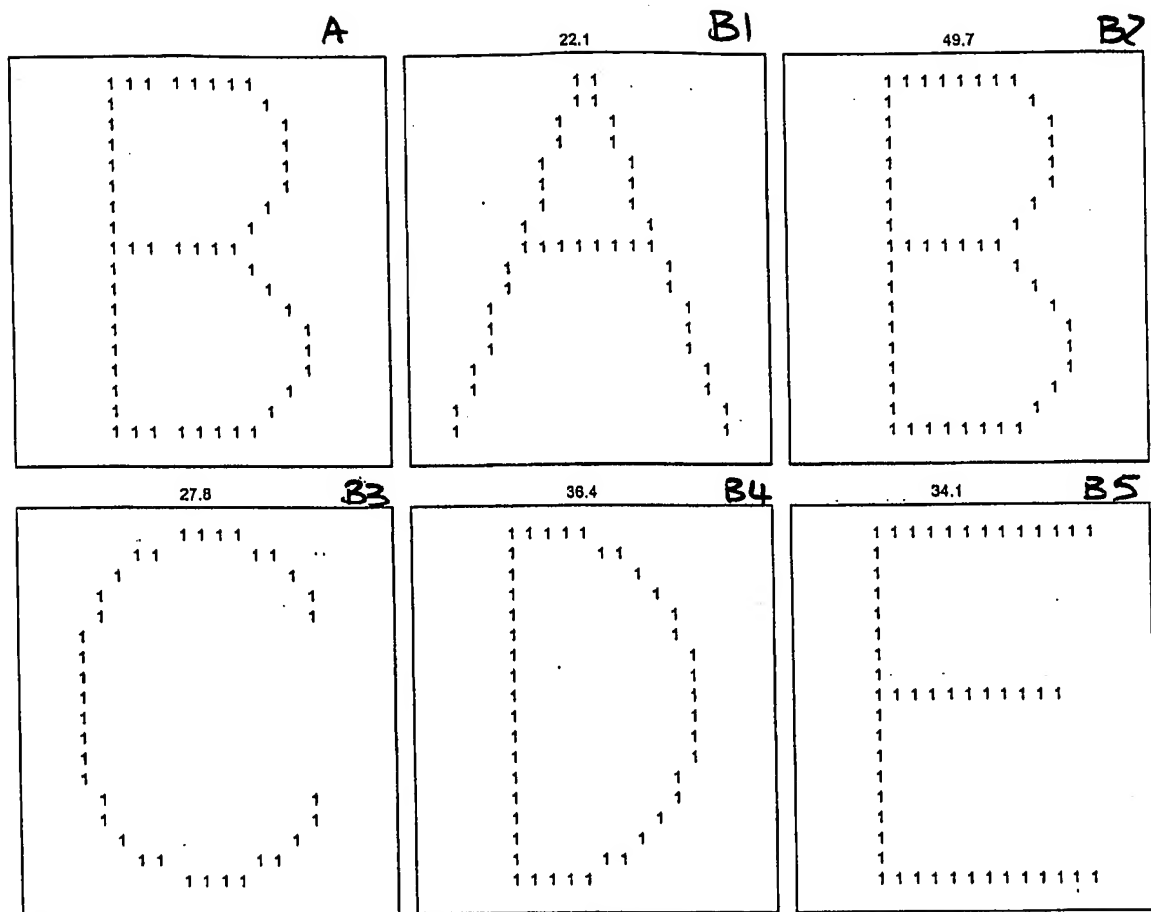


Figure 4

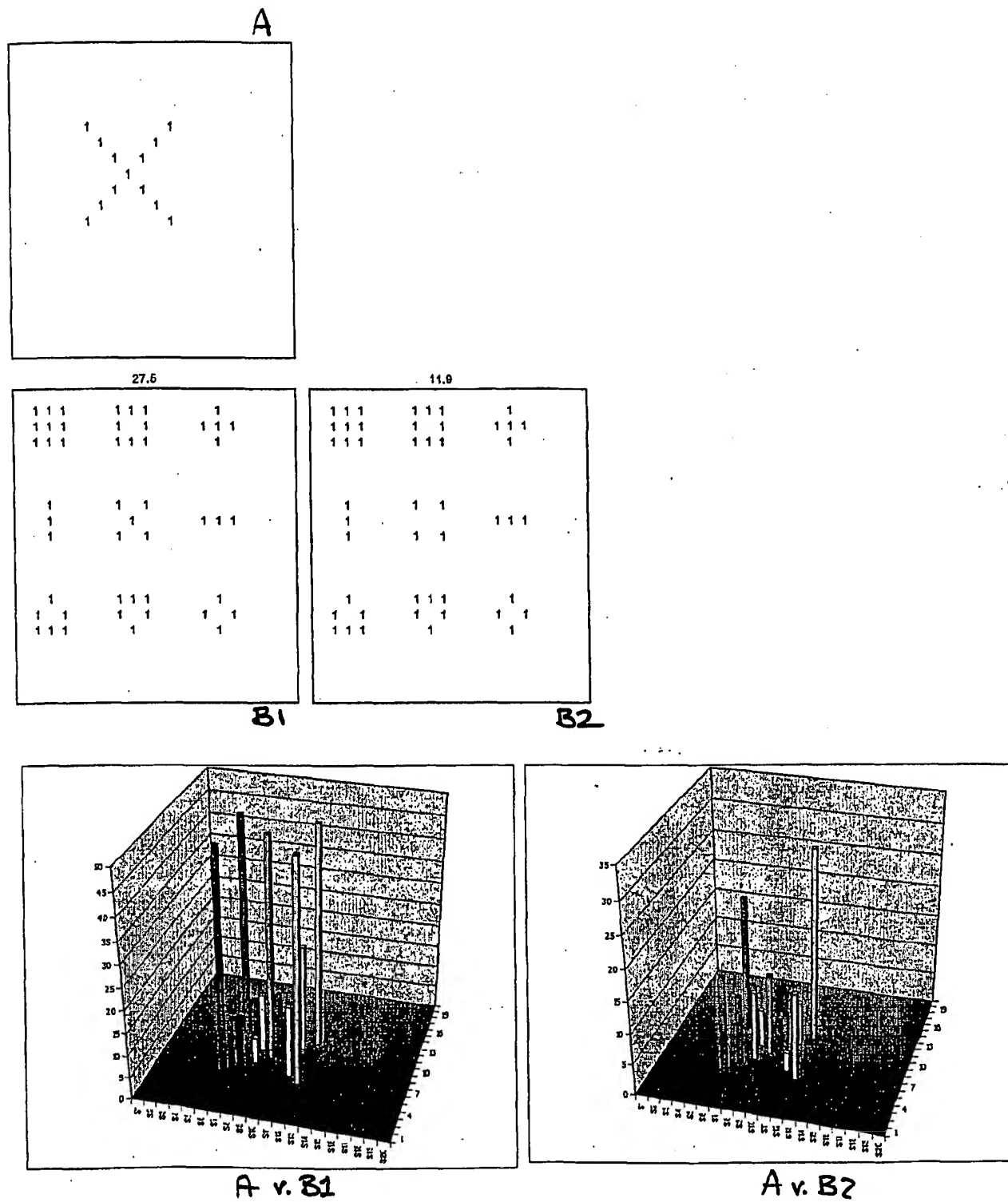
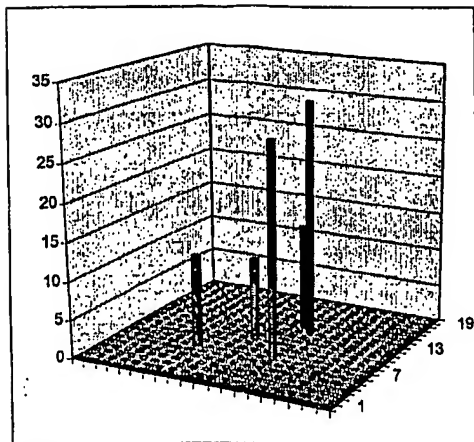
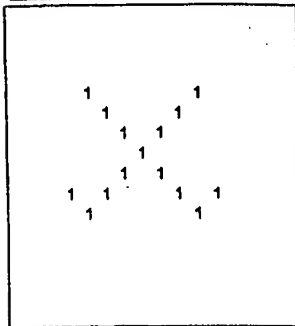


Figure 5



A v. B4



A

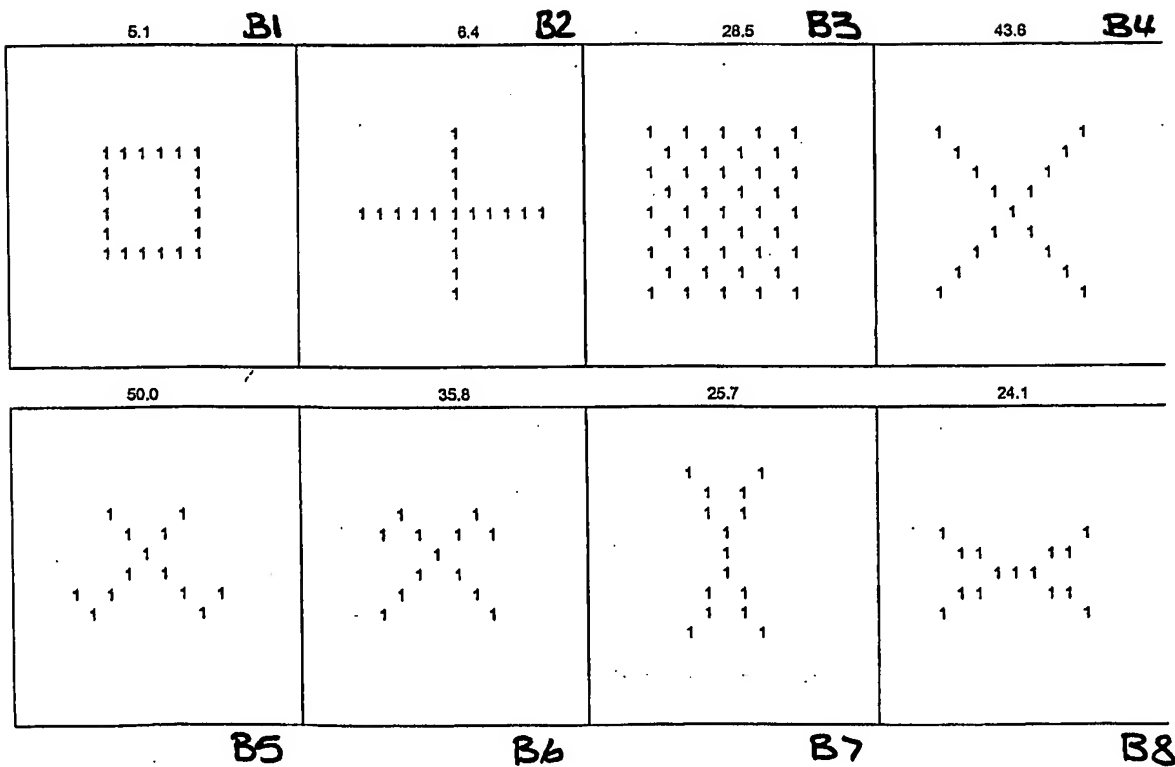


Figure 6

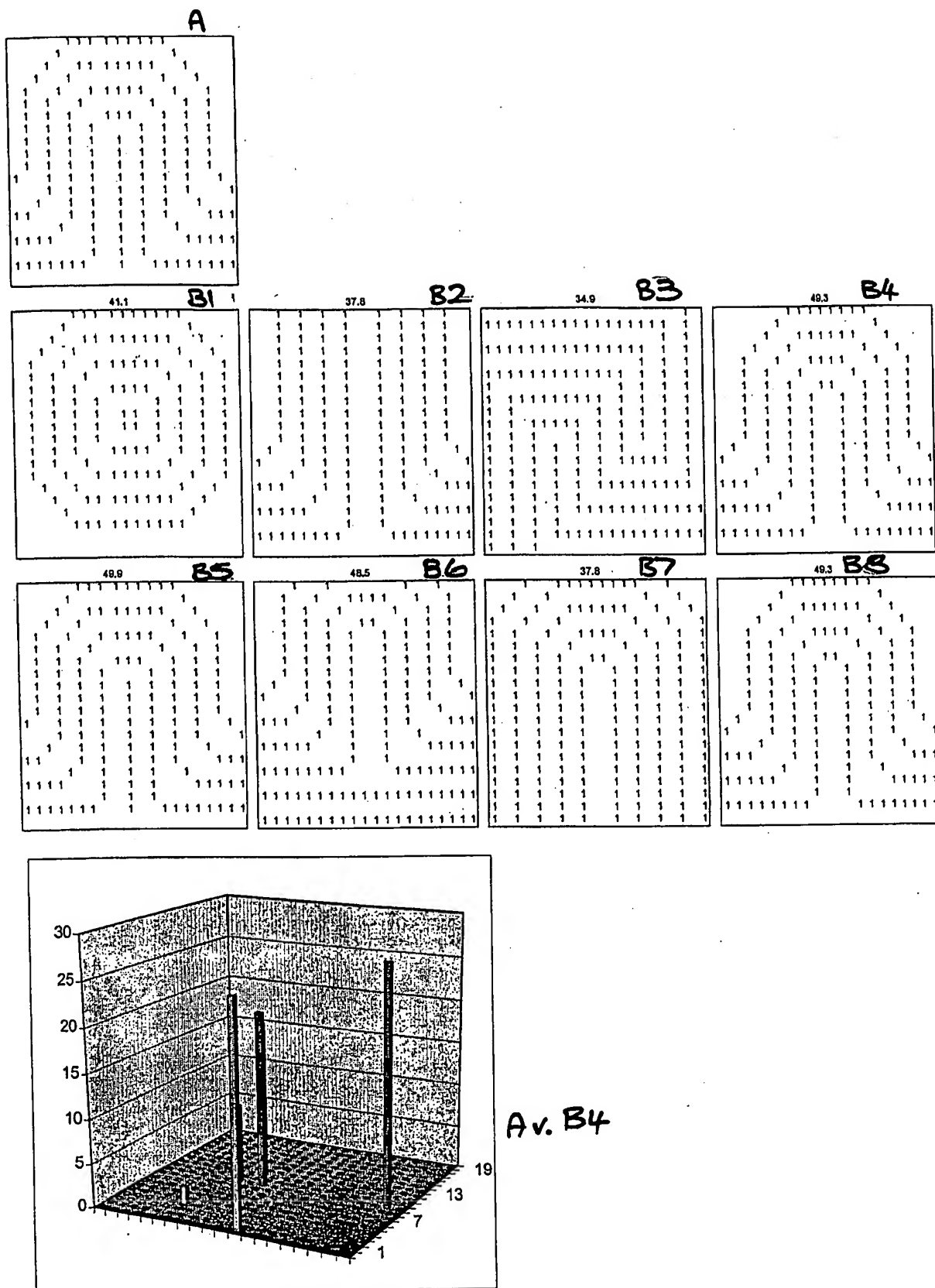


Figure 7

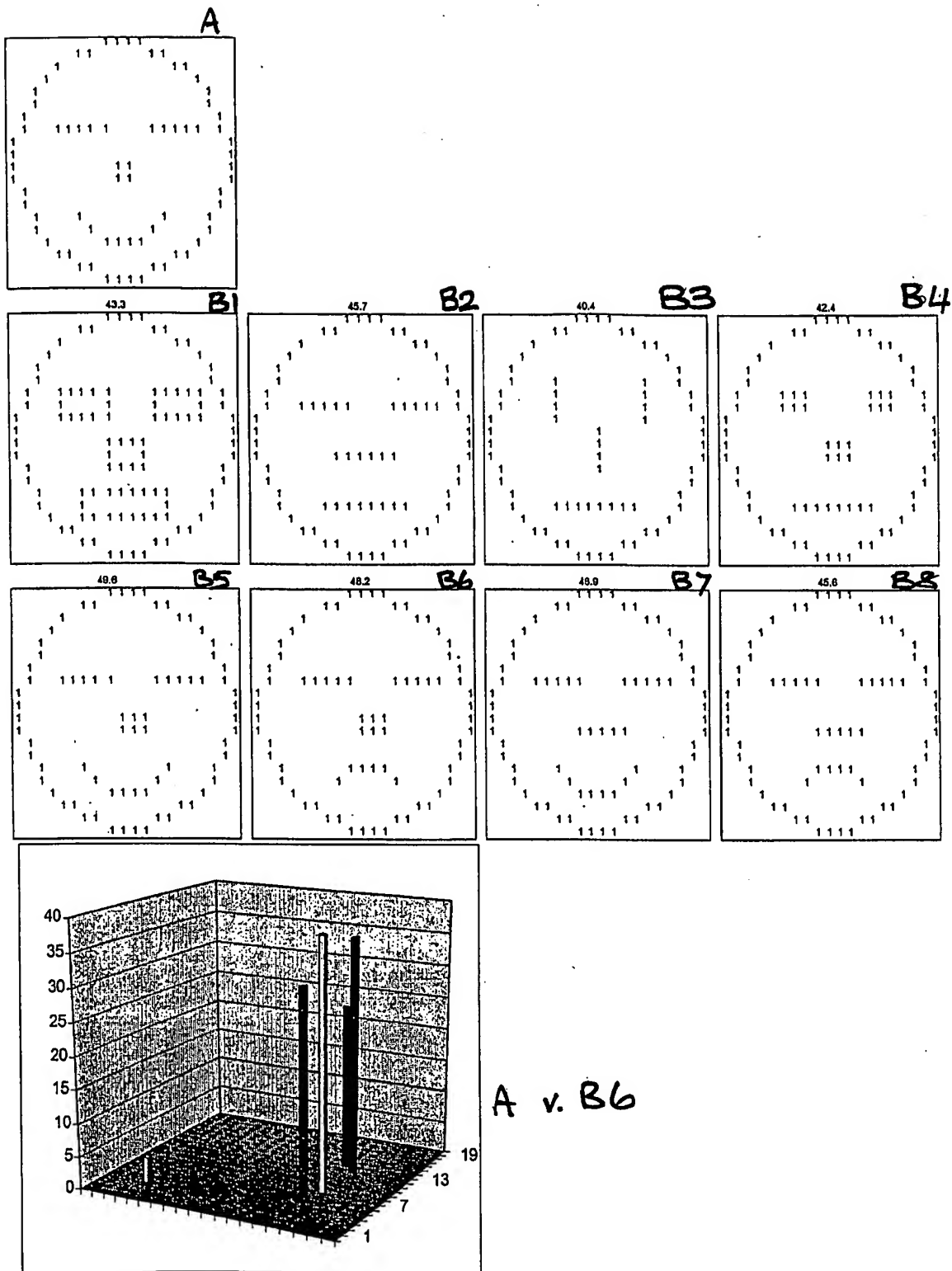


Figure 8

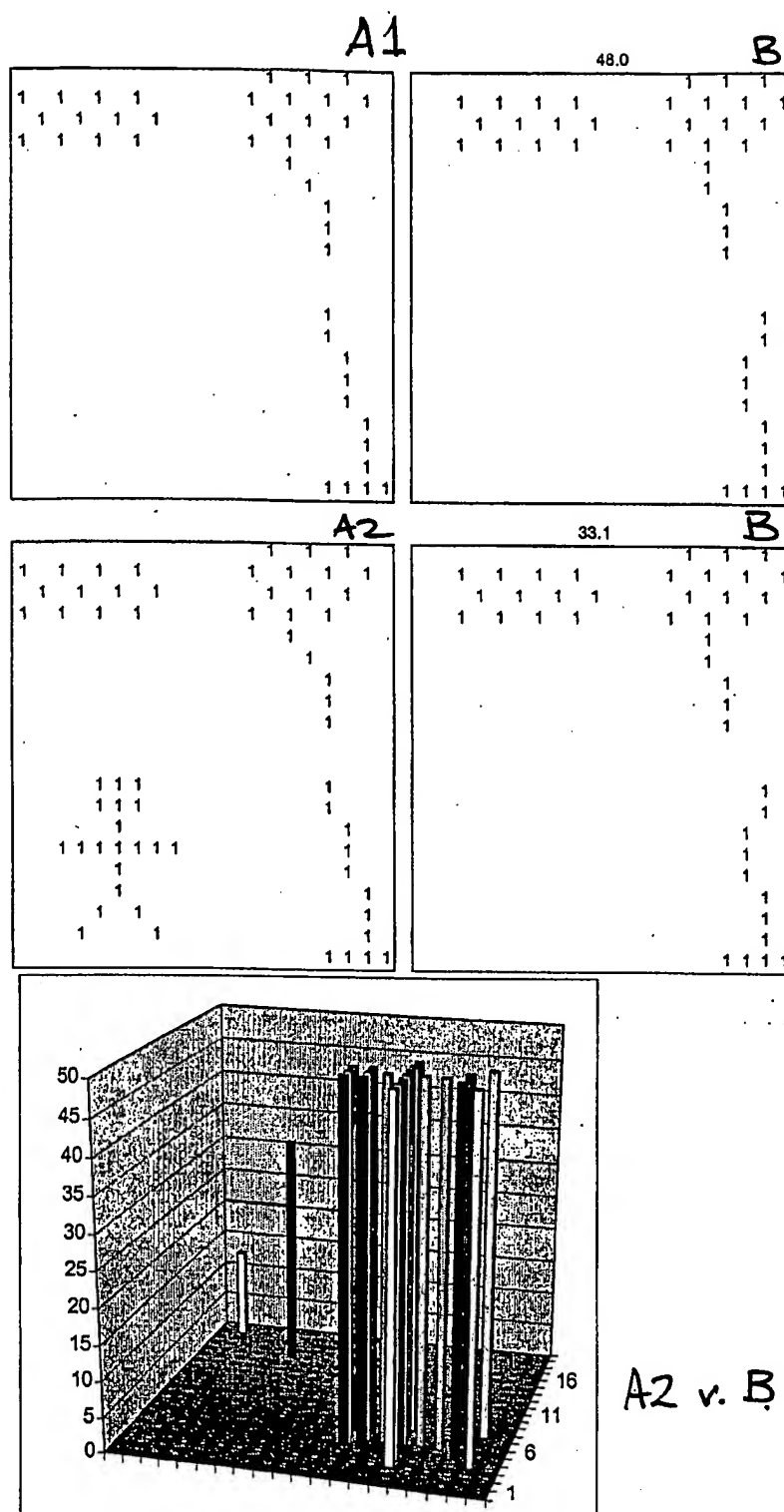


Figure 9

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/01209

A. CLASSIFICATION OF SUBJECT MATTER		
IPC 7	G06K9/64	G06K9/00 G06T7/00
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
IPC 7 G06K G06T		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 1 126 411 A (BRITISH TELECOMM) 22 August 2001 (2001-08-22) cited in the application abstract; figure 1 page 4, line 3 -page 5, line 39	1-10
A	STENTIFORD F W M: "An Evolutionary Programming Approach to the Simulation of Visual Attention" CONGRESS ON EVOLUTIONARY COMPUTATION, 27 - 30 May 2001, pages 851-858, XP001098221 Seoul, Korea cited in the application abstract; figures 3A,3B page 852 -page 853 -/-	1-10
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *Z* document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
27 May 2003		10/06/2003
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Müller, M

INTERNATIONAL SEARCH REPORT

Internat Application No

PCT/GB 03/01209

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ROHWER R ET AL: "The Theoretical and Experimental Status of the n-tuple Classifier" NEURAL NETWORKS, ELSEVIER SCIENCE PUBLISHERS, BARKING, GB, vol. 11, no. 1, 1998, pages 1-14, XP004109175 ISSN: 0893-6080 abstract page 2, left-hand column, paragraph 2 -page 3, left-hand column, paragraph 3 -----	1-10
A	GALLET O ET AL: "A model of the visual attention to speed up image analysis" IMAGE PROCESSING, 1998. ICIP 98. PROCEEDINGS. 1998 INTERNATIONAL CONFERENCE ON CHICAGO, IL, USA 4-7 OCT. 1998, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 4 October 1998 (1998-10-04), pages 246-250, XP010308736 ISBN: 0-8186-8821-1 -----	

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

ation on patent family members

International Application No

PCT/GB 03/01209

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 1126411	A	22-08-2001	EP 1126411 A1	22-08-2001
			AU 3202901 A	27-08-2001
			CA 2400085 A1	23-08-2001
			EP 1281164 A2	05-02-2003
			WO 0161648 A2	23-08-2001
			US 2002081033 A1	27-06-2002
